

Genforce version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2001, 14:29:18 : Search time 58.44 Seconds  
(without alignments)  
216.295 Million cell updates/sec

Title: US-09-593-316-2

Perfect score: 2003  
Sequence: 1 MNKGRVLSMLVSTIVV.....IKLVKMSOTREYVNNV 369

Scoring table: BL0SDM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 3425486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1906.5	95.2	368	1	GATR_BOVIN
2	1711	85.4	371	1	GATR_PIG
3	1471	73.4	394	1	GATR_MOUSE
4	670	33.4	354	1	BCAT_HUMAN
5	110	5.5	505	1	AGAR_ALTAT
6	96	4.8	583	1	PR40_YEAST
7	94.5	4.7	408	1	Y233_METJA
8	94.5	4.7	563	1	YMN0_YEAST
9	94.5	4.7	688	1	TALA_POVIC
10	93	4.6	1057	1	POL_STVAT
11	91.5	4.6	522	1	HEX1_ENTHI
12	89.5	4.5	959	1	DPO1_AERPE
13	89	4.4	659	1	HS82_ORYSA
14	88.5	4.4	517	1	DMPN_FESFP
15	87	4.3	536	1	VHT1_HET7
16	87	4.3	569	1	GTL1_DROME
17	86.5	4.3	510	1	HVAL_MACEA
18	86.5	4.3	691	1	TALA_POVHA
19	86.5	4.3	733	1	YLC4_CABEL
20	86.5	4.3	2410	1	POLL_BAYMJ
21	85.5	4.3	708	1	TALA_SVA0
22	85	4.2	863	1	PHSC_MYCTU
23	85	4.2	1276	1	BXD_CLOHO
24	84.5	4.2	185	1	VOJM_YEAST
25	84.5	4.2	514	1	THRC_SCHPO
26	84.5	4.2	558	1	0166_BACCE
27	84.5	4.2	2386	1	RAD3_SCHPO
28	84	4.2	408	1	ARCA_MYCHO
29	84	4.2	446	1	GLNA_METJP
30	84	4.2	727	1	IF2M_BOVIN
31	84	4.2	1134	1	YML7_YEAST
32	83.5	4.2	434	1	KP58_MOUSE
33	83.5	4.2	461	1	SHL1_YEAST

34	83.5	4.2	467	1	IRF6_HUMAN	014896 homo sapien
35	83.5	4.2	502	1	MTFB_AZOVI	P11067 azotobacter
36	83	4.1	355	1	Y951_MEOTA	Q58361 methanococ
37	83	4.1	639	1	GLOB_HUTFI	P10539 butylalbu
38	83	4.1	992	1	VP41_YEAST	P48959 saccharomy
39	83	4.1	1451	1	SPT6_YEAST	P36615 saccharomy
40	82.5	4.1	368	1	MNN9_CANAL	P36697 candida alb
41	82.5	4.1	436	1	KP58_HUMAN	P21127 homo sapien
42	82.5	4.1	1586	1	SNZ2_HUMAN	P15141 homo sapien
43	82	4.1	454	1	GLNA_METJA	Q60182 methanococ
44	82	4.1	559	1	IPAGI_BOVIN	Q07547 bos taurus
45	82	4.1	699	1	HS80_LYCDS	P61811 lycopersico

## ALIGNMENTS

RESULT 1	
GATR_BOVIN	STANDARD: PRT: 368 AA.
ID GATR_BOVIN	
AC P14769:	
DF 01-APR-1990 (Rel. 14, Created)	
DT 01-APR-1990 (Rel. 14, Last sequence update)	
DT 30-MAY-2000 (Rel. 39, Last annotation update)	
DE N-ACETYLTRANSFERASE (UDP-GALACTOSE + BETA-D-GALACTOSYL-1,4-GLUCOSAMINE) (EC 2.4.1.15)	
DE (GALACTOSYLTRANSFERASE) (UDP-GALACTOSE + BETA-D-GALACTOSYL-1,4-N-ACETYL-D-GALACTOSAMINE) (EC 2.4.1.15)	
GN GATR1	
OS Bos taurus (bovine)	
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.	
OX NCBI_TaxID=9913;	
KN [1]	
RP SEQUENCE FROM N.A.	
RX MEDLINE:8940543; PubMed:2503516;	
RA Joziasse D.H., Shaper J.H., van den Eijnden D.H., van Tunen A.J., Shaper N.H.	
RT "Bovine alpha 1->3-galactosyltransferase: isolation and characterization of a cDNA clone. Identification of homologous sequences in human genomic DNA."	
RL J. Biol. Chem. 264:14290-14297(1989).	
CC -1- FUNCTION: TRANSFER OF GALACTOSE FROM UDP-GALACTOSE TO AN ACCEPTOR MOLECULE (R).	
CC -1- CATALYTIC ACTIVITY: UDP-GALACTOSE + BETA-D-GALACTOSYL-1,4-N-ACETYL-D-GALACTOSAMINE-R = UDP + ALPHA-D-GALACTOSYL-1,3-BETA-D-GALACTOSYL-1,4-N-ACETYL-D-GALACTOSAMINE-R.	
CC -1- COFACTOR: MANGANESE.	
CC -1- PATHWAY: GLYCOSYLATION.	
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND FORN IN TRANS CISTERNAE OF GOLGI.	
CC -1- DISEASE: AUTOIMMUNE DISEASE (ANTIBODIES AGAINST ENZYMAIC PRODUCT).	
CC -1- SIMILARITY: STRONG, TO BLOOD GROUP AB TRANSFERASE.	
CC -1- SIMILARITY: STRUCTURAL SIMILARITY WITH THE OTHER MAMMALIAN GLYCOSYLTRANSFERASES.	
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announc/">http://www.isb-sib.ch/announc/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC EMBL: J04989; AAA30558.1; -	
DR PIR: A44785; A44785.	
KW Signalase: Glycosyltransferase; Glycoprotein; Transmembrane; Signal-anchor; Golgi stack; Manganese.	
KM Transmembrane (POTENTIAL).	
FT DOMAIN 1 6 CYTOPLASMIC (POTENTIAL).	
FT TRANSMEM 7 22 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).	
FT DOMAIN 23 368 LUMENAL, CATALYTIC (POTENTIAL).	
FT CARBOHYD 293 293 N-LINKED (GLCNAC... ) (POTENTIAL).	



```

CC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX      MBL_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=90046769; PubMed=2510162;
RA      Larsen R.D., Rajan V.P., Ruit M.M., Kukowska-Latallo J.,
RA      Cummings R.D., Lowe J.B.;
RT      "Isolation of a cDNA encoding a murine
RT      UDPgalactose-beta-D-galactosyl-1,4-N-acetyl-D-glucosaminide
RT      alpha-1,3-galactosyltransferase: expression cloning by gene
RT      transfer."
RL      Proc. Natl. Acad. Sci. U.S.A. 86:8227-8231(1989).
RL      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=92184813; PubMed=1544928;
RA      Joiasse D.H., Shaper N.L., Kim D., van den Eijnden D.H., Shaper J.H.;
RT      "Murine alpha 1,3-galactosyltransferase. A single gene locus
RT      specifies four isoforms of the enzyme by alternative splicing."
RL      J. Biol. Chem. 267:5534-5541(1992)
CC      -1- FUNCTION: TRANSFER OF GALACTOSE FROM UDP-GALACTOSE TO AN
CC      ACCEPTOR MOLECULE (R).
CC      -1- CATALYTIC ACTIVITY: UDP-GALACTOSE + BETA-D-GALACTOSYL-1,4-
CC      N-ACETYL-D-GLUCOSAMINYL-R -> UDP + ALPHA-D-GALACTOSYL-1,3-
CC      BETA-D-GALACTOSYL-1,4-N-ACETYL-D-GLUCOSAMINYL-R.
CC      -1- COFACTOR: MANGANESE.
CC      -1- PATHWAY: GLYCOSYLATION.
CC      -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
CC      FORM IN TRANS EISTERNAE OF GOLGI.
CC      -1- SIMILARITY: STRONG, TO BLOOD GROUP AB TRANSFERRIN.
CC      -1- SIMILARITY: STRUCTURAL SIMILARITY WITH THE OTHER MAMMALIAN
CC      GLYCOSYLTRANSFERASES.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to licensing@sib-sib.ch).
CC      -----
DR      EMBL; M26925; AAA37657.1; -;
DR      EMBL; M85153; AAA37711.1; -;
DR      PIR; A34417; A14417.
DR      MGD; MGI:95704; Atrial.
KW      Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW      Signal-anchor; Golgi stack; Alternative splicing; Manganese.
FT      DOMAIN 1 41
FT      TRANSMEM 42 60
FT      DOMAIN 61 394
FT      CARBOHYD 83 83
FT      CARBOHYD 319 319
FT      VARSPLIC 1 35
FT      VARSPLIC 62 62
FT      SEQUENCE 394 AA: 46475 MW: 7766816kDIDIBBP CQC64;

```

```

Query Match 36: 73.4% Score 1471: DB 1: Length 394:
Best Local Similarity 71.5%: Pred. No. 9e-112:
Matches 266: Conservative 47: Mismatches 43: Indels 16: Gaps 3:

QY 1 MNNKGVKIIISLVSVIVFVMEVYIHSPEGSILFWITNPNSNPVSGSSSTOKGGMFPRFN 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 36 MNNKGVKIIISLVSVIVFVMEVY-----NKIPEV-GENNRQKQKQKMFSPSK 82
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 61 NG---YQEDHDVDPEKKCKPRKFNKSKLKLSDMFPNFKRPREVITMDKAPVYWGSTYNBA 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 83 NGTSHYQHDNVAQGRKKRKNDRIDIEEPQILMDFNFKRNPDPVLTVPKAPITWGSTYDTA 142
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 118 VDDYFAAKITVGLTTFVAVRFRFYTHYILEPFLTSNKHKHFMRHRYLFVYMDVDSRMLPI 177
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 143 ELEKTYATVKKLVGLTTFVAVCRITLHVTLEDFLEKSDMTFMGCHRIYLFVYMDVDSRMLPV 202
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

QY	178	ELGRLSRKPYFVYKPRPKRWQDVMYMKRTTIGSHVANHIGREDELFCMDYDQVYVQDDPQV	247
Db	203	HLNPLSLQVLEFVLRSEKRWQDISMKMRKRTTIGHLIHLIHGFEDFLFCMDYDQVVDQNDG	262
QY	238	ETLSESVAGVQVMWYKADPEEFYERKSKSAVTPGEGRFVYHAIKGGPTGVLTNTQ	297
Db	263	ETLSQLVAQVQVMWYKASPKFTYERKRELSAATPGEGRFVYHAIKGGPTHTLTNTLR	322
QY	298	ECEFGILAKDKRNDIEAGWDHSHLNKYPFLINKPTKTLSPDYGMVHLSIPADIKLVKSM	357
Db	323	ECEFGILQDKRNDIEAGWDHSHLNKYPFLINKPTKTLSPDYGMVHLSIPADIKLVKSM	382
QY	358	QTRKFNVAKNV	369
Db	383	QTRKFNVAKNV	394
RESULT 4			
HGAT_HUMAN			
ID	BGAT_HUMAN	STANDARD:	PRF: 354 AA.
AC	P16442:		
DT	01-AUG-1990 (Ref. 15, Created)		
DI	01-DEC-1992 (Ref. 24, Last sequence update)		
DT	01-OCT-2000 (Ref. 40, Last annotation update)		
DE	HISTO-BLOOD GROUP ABO SYSTEM TRANSFERASE (NAGAT) [INCLUDES:		
DE	GLYCOPROTEIN-FUCOSYLTRANSFERASE (EC 2.4.1.40) (FUCOSYLGLYCOPROTEIN		
DE	ACTIVATINGALACTOSAMINYLTANSFERASE (EC 2.4.1.44) (GLYCOSYLALACTOSAMINYLTANSFERASE)		
DE	ALPHA-N-ACETYLALACTOSAMINYLTANSFERASE) (HISTO-BLOOD GROUP A		
DE	TRANSFERRASE) (A TRANSFERRASE); GLYCOPROTEIN-FUCOSYLALACTOSIDE ALPHA-		
DE	GALACTOSYLTRANSFERASE (EC 2.4.1.37) (FUCOSYLGLYCOPROTEIN + ALPHA-		
DE	GALACTOSYLTRANSFERASE) (HISTO-BLOOD GROUP B TRANSFERRASE) (B		
DE	TRANSFERRASE)]].		
GN	Homo sapiens (human).		
OS	Homo sapiens (human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.		
OX	NCBI_Taxid:9606;		
RN	[1]		
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.		
RX	MEDLINE:90110098; PubMed:2104828;		
RA	Yamamoto F.-I., Marken J., Tsuji T., White T., Clausen H.,		
RA	Hakomori S.-I.;		
RT	"Cloning and characterization of DNA complementary to human Hb-		
RT	GalNAc: Fuc alpha 1->2Gal alpha 1->3GalNAc transferase (histo-blood		
RT	group A transferase) mRNA.";		
RL	J. Biol. Chem. 265:1146-1151(1990).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE:90238543; PubMed:2333095;		
RA	Yamamoto F.-I., Clausen H., White T., Marken J., Hakomori S.-I.;		
RT	"Molecular genetic basis of the histo-blood group ABO system.";		
RL	Nature 345:229-233(1990).		
RN	[3]		
RP	CHARACTERIZATION.		
RX	MEDLINE:91035463; PubMed:2121736;		
RA	Yamamoto F.-I., Hakomori S.-I.;		
RT	"Sugar-nucleotide donor specificity of histo-blood group A and B		
RT	transferases is based on amino acid substitutions.";		
RL	J. Biol. Chem. 265:19257-19262(1990).		
CC	-1- FUNCTION: THIS PROTEIN IS THE BASIS OF THE ABO BLOOD GROUP SYSTEM.		
CC	THE HISTO-BLOOD GROUP AB INVOLVES THREE CARBOHYDRATE ANTIGENS: A,		
CC	B, AND H. A, B, AND AB INDIVIDUALS EXPRESS A GLYCOSYLTRANSFERASE		
CC	ACTIVITY THAT CONVERTS THE H ANTIGEN TO THE A ANTIGEN (BY ADDITION		
CC	OF UDP-GALNAc) OR TO THE B ANTIGEN (BY ADDITION OF UDP-GAL).		
CC	WHEREAS O INDIVIDUALS LACK SUCH ACTIVITY.		
CC	-1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GALACTOSAMINE + GLYCOPROTEIN		
CC	ALPHA-D-FUCOSYL-(1,2)-D-GALACTOSE -> UDP + N-ACETYL-ALPHA-D-		
CC	GALACTOSAMINYL-(1,3)-[ALPHA-D-FUCOSYL-(1,2)]-D-GALACTOSE.		
CC	-1- CATALYTIC ACTIVITY: UDP-GALACTOSE + GLYCOPROTEIN ALPHA-D-FUCOSYL-		
CC	(1,2)-D-GALACTOSE -> UDP + GLYCOPROTEIN ALPHA-D-GALACTOSYL-(1,4)-		
CC	[ALPHA-D-FUCOSYL-(1,2)]-D-GALACTOSE.		
CC	-1- PATHWAY: GLYCOSYLATION.		
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN; MEMBRANE-BOUND		



QY 224 CNDVDFE-----DERGVETIGESV-----AQLQAWW- 252  
 DB 286 -LEDDNREFKSRLEATLDMIDVSLALGETKRVISEYSGRRRPNENAPWALRKDWFL 344  
 QY 253 -KAUDEFYERKRESA-----AVIFEGEDFYHAALFG-----GTFYVAINLTQCEFCGI 303  
 DB 345 KIASPMLMQLFSLSPDVISLIFVPI-----KALKGTADADGTYYNKKILLRQ----- 391  
 QY 304 LKQKNDLEAOW 315  
 DB 392 -KEAPNCTGENW 402

RESULT 6  
 PRD\_YEAST STANDARD; PRT: 583 AA.  
 ID PR40\_YEAST STANDARD; PRT: 583 AA.  
 AC P33203;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE PRE-MRNA PROCESSING PROTEIN PRP40.  
 GN PRP40 OR YKL012W OR YKL165.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN 111  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93127732; PubMed=1481574;  
 RA Pascolo S., Ghazvini M., Boyer J., Colliaux L., Thierry A., Dujon B.;  
 RT "The sequence of a 9.3 kb segment located on the left arm of the  
 RT yeast chromosome XI reveals five open reading frames including the  
 RT CCE1 gene and putative products related to MYO2 and to the ribosomal  
 RT protein L10.\*";  
 RL Yeast 8:987-995(1992).  
 RN 121  
 RP IDENTIFICATION  
 RX MEDLINE=96182100; PubMed=8622649;  
 RA Kao H.-Y., Siliciano P.G.;  
 RT "Identification of Prp40, a novel essential yeast splicing factor  
 RT associated with the U1 small nuclear ribonucleoprotein particle\*";  
 RL Mol. Cell. Biol. 16:960-967(1996).  
 RN 131  
 RP PARTIAL SEQUENCE  
 RX MEDLINE=97165029; PubMed=9012791;  
 RA Neubauer G., Gottschalk A., Fabrizio P., Seraphin B., Luenemann R.,  
 RA Mann M.;  
 RT "Identification of the proteins of the yeast U1 small nuclear  
 RT ribonucleoprotein complex by mass spectrometry\*";  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:385-390(1997).  
 CC -1- FUNCTION: REQUIRED FOR THE FIRST STEP OF PRE-MRNA SPLICING. THIS  
 CC PROTEIN IS ASSOCIATED WITH SNRP U1.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- SIMILARITY: CONTAINS 2 WW DOMAINS.  
 CC -1- SIMILARITY: WEAK SIMILARITY TO YEAST MYO2.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: S53418; AAB24902.1; -  
 DR EMBL: Z28012; CAAB1847.1; -  
 DR PIR: S30014; S30014.  
 DR SDD: S0001495; PRP40.  
 DR InterPro: IPR001202; -  
 DR InterPro: IPR002713; -  
 DR Pfam: PF01846; PF: 4.  
 DR Pfam: PF00197; WW: 2.

DR PROSITE: PS01159; WW DOMAIN\_1; 2.  
 DR PROSITE: PS00920; WW DOMAIN\_2; 2.  
 KM mRNA processing; mRNA splicing; Nuclear protein; Ribonucleoprotein;  
 KM Repeat.  
 FT DOMAIN 1 31 WW.  
 FT DOMAIN 2 72 WW.  
 FT MUTAGEN 240 240 S->P; IN PRP40 1 SUPPRESSOR; AFFECTS SAK1  
 FT mRNA ACCUMULATION IN U1-U4 MUTANT AT 18  
 FT DEGREES CELSIUS.  
 SU SEQUENCE 583 AA; 69065 MW; BAICIGYID532924C CPG64;

Query Match 4.88; Score 96; In 1; Length 583;  
 Best Local Similarity 20.98; Prod No. 2.5;  
 Matches 67; Conservative 51; Mismatches 101; Indels 102; Gaps 20;

QY 67 DEDVEERKREKRSKLT--KLSDFNPKREVVMTQKAPVWECTYNNKAVLDYYA 124  
 DB 256 DILDTQKSKKKKLTQALKELREYLN-----GIITSSSEFTITWQILNHYVD----- 306  
 QY 125 KQRTVGLTFVAVGRITYEELFPLSANKRPV-GHRYT---FVYVDDVSRLMLELG 180  
 DB 307 KSK-----RYM-----ANRRKVLTHEVLNEYLKIYNTENDLQNNLIN 345  
 QY 181 PLR-----SFK-----VFVYPERMDVSVRRKTTGEMIVAHITQREVDPL 222  
 DB 346 ELRLRYTDRARDNFKSLREVPKIKANTKMSD-----IYPIKSDPRFL 393  
 QY 223 F-----CNDVDFE-----FGVETIGESVAQLQAW WYKADDEFYERKRESA 268  
 DB 394 HMLGRGSSCLDLPLDFVDEQMYIFAOKSLAQYLLDKQNFPMADVSDITKQNIKLV 453  
 QY 269 AVIFEGEDFYHAALFGSTPTVNLITQCEFGKILKCKKNIDKQNDHPSLL-----NKY 324  
 DB 454 -----END-----RKFDVDEKDESLVD---GLIKRNEKIQOKLQKLEDEKKRY 499  
 QY 325 F--LUNKP--FKLLSP--YQWD 341  
 DB 500 FWLLDQRTYTKIGKRPSTWD 520

RESULT 7  
 Y233\_METJUA STANDARD; PRT: 408 AA.  
 ID Y233\_METJUA STANDARD; PRT: 408 AA.  
 AC Q60290;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE HYPOTHETICAL PROTEIN MDEC133.  
 GN MDEC133.  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;  
 OC Methanococcus.  
 OX NCBI\_TaxID=2190.  
 RN 111  
 RP SEQUENCE FROM N.A.  
 RP STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;  
 RA Bull C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald J.M., Clayton R.A., Gocayne J.D.,  
 RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Keich C.L.,  
 RA Overback R., Kirkness E.P., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Gishagen N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,  
 RA Ulteback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Clifton H.-P., Fraser C.M., Smith H.O., Moeese C.R., Venter J.C.;  
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 RT jannaschii.\*";  
 RL Science 273:1058-1073(1996).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its





use by non-profit institutions as long as its content is in no way modified and this statement is not removed, usage by and for commercial entities requires a license agreement (See <http://www.isb-stb.ch/announce/> or send an email to [license@isb-stb.ch](mailto:license@isb-stb.ch)).

EMBL: 009759; AA00165.1; ...  
 DIR HSB2: 106665; 106665;  
 DIR InterPro: IP001540;  
 DIR Pfam: PF00728; Glyco hydro 20; 1;  
 DIR PRINTS: PR00740; GLYDOLASE20;  
 KM Hydrolyses glycylase;  
 SV SEQUENCE: 522 AA; 594 to MM; 1793034565683AB GRC64;

Query Match 4.98; Score 91.5; DB 1; Length 522;  
 Best Local Similarity 20.8%; Prod. No. 5;  
 Matches 58; Conservativ 44; Mismatches 84; Indels 95; Gaps 16;  
 Y 17 VIVVWETYSRPE --- GSLEWNPISRP --- EWS --- GSSICK 91  
 D 246 VVANWVWVIVSVSMKRGVNLSTNPA-NRTPPIIDALMKELSDTGTDTYVHGGEV --- 402  
 Y 52 GWVFPWVFNKVEDEEDVDKPKPKSKL KLSMNPDKREVVPTTKAVVW 110  
 D 403 --- WISWMSKSKESVSLQKPKSKSLNSLDELYFNKACQVLI-HICKHVVW 353  
 Y 111 EETYNKA --- VVVVYAKOKITV -GVFAVGYTHYLE --- EETTSANK 154  
 D 504 EEWKKNALKNITLDVWOKRLQVNSGKATPSASHTYAKQMPILNSYSSSTVNT 413  
 Y 155 HENVHIVLEVYVW --- WWSRMPLE --- KSKVE 188  
 D 414 HSM --- WVVFNIDMYDINPVKSSSEKENVLGGKSSMSTEDKQNFDPVQHY 466  
 Y 189 EVKTEHGMVSWMRKT IGEHIVAHVREVDLEMDV 227  
 D 467 SATAEKSKSVYKES --- IRY --- KANTLRIDV 497

RESULT 12  
 Description: STANDARDS; PRT: 959 AA.  
 ID ID01: AERUE  
 AC 093745;  
 DT 01-MAY-2000 (rel. 49; Created)  
 DT 01-MAY-2000 (rel. 49; Last sequence update)  
 DT 01-MAY-2000 (rel. 49; Last annotation update)  
 DB DNASIS: 1 (rel. 2.7.1.7).  
 DE POLA OR APPE099;  
 GN Acetylglutamate;  
 OS Acetobacter glaucus (Acetobacter); Acetobacteraceae;  
 AC Acetobacter;  
 CC NCBI Taxid 56636;  
 RN [1]  
 RP SEQUENCE OF 47 959 FROM N.A.  
 RC STRAIN-K1;  
 RA Ishio Y., Oami I.K.;  
 RT "Isolation of the genes encoding two alpha-like DNA polymerases from Acetobacter peritumescens".  
 RL Submitted (Sep-1998) to the EMBL/Genbank/JDBI databases.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K1;  
 RA MEDLINE: 9941049; PubMed: 10492966;  
 RA Kawaiyoshi Y., Hino Y., Horioka H., Yamazaki S., Hakiwa Y., Aino H., Takahashi M., Sakino M., Baba S.-I., Anai A., Kosugi H., Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H., Takamiya M., Masuda S., Funahashi T., Tanaka T., Kogoh Y., Yamazaki J., Koshida N., Gotoh A., Aoki K.-I., Kubota K., Nakamura Y., Minoda N., Sako Y., Kikuchi H.;  
 RA "Complete genome sequence of an acetate hyper-tolerant *Acetobacter peritumescens* strain K1".  
 RT GenBank: 101 (1999).  
 RL 1. ANALYTIC ACTIVITY: N-DI-OXYMETHYLPHOSPHATE TRIPHOSPHATE - N

CC PYROPHOSPHATE + DNA(N).  
 CC 1. SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.

CC This SWISS-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation. The European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed, usage by and for commercial entities requires a license agreement (See <http://www.isb-stb.ch/announce/> or send an email to [license@isb-stb.ch](mailto:license@isb-stb.ch)).

EMBL: A017500; HAAT5662.1; ...  
 DIR EMBL: A000058; HAAT9008.1; ...  
 DIR InterPro: IP002404; ...  
 DIR Pfam: PF00136; DNA pol B; 2;  
 DIR PRINTS: PR00106; DNA POLYMERASE-B; 1;  
 DIR PROSITE: PS00106; DNA-POLYMERASE-B; 1;  
 KM Transferrase; DNA-directed DNA polymerase; DNA replication;  
 D DNA-binding;  
 FT CONFLICT 193 193 L -- H (IN REF. 1).  
 FT CONFLICT 456 462 R -- G (IN REF. 1).  
 FT CONFLICT 552 552 R -- G (IN REF. 1).  
 FT CONFLICT 555 555 L -- V (IN REF. 1).  
 FT CONFLICT 822 822 K -- R (IN REF. 1).  
 SV SEQUENCE 959 AA; 109748 MM; 0546636908885 GRC64;

Query Match 4.98; Score 89.5; DB 1; Length 959;  
 Best Local Similarity 22.4%; Prod. No. 16;  
 Matches 58; Conservativ 28; Mismatches 81; Indels 95; Gaps 11;

Y 158 VQHRV -EVVWVIVSRKMLLEGLSKPKVEK --- PPKQVIVSWVNRK 205  
 D 633 VQHRVKEKQSNEL --- VMLDKFVRLYKRSKSKSLREERKMTDVVSANK 685  
 Y 206 IGEHIVAHVREVDLEMDVQVDFEVEVETLGSVAQIDQWVAKADPEVTEERK 265  
 D 686 VVNSVGVFGSKSLVSLVAVR --- SVTALHVALR --- GILKSKR 727  
 Y 266 ESNATVPG --- EDPYHAAEKGCTVYVNI 295  
 D 728 ELMHLYVDQSTLPLMDPKQVINDVYVETVGLLELAKVRAVILSSLSKKNYLT 787  
 Y 296 TQF --- CFFGLDKKNDLEAQMDESHLNKYL --- INKPKIISPEYCWVHGLAD 349  
 D 788 TEGDQVLEKQVAKKSNTE --- FIKDFSKAVKILS --- KLEKED 828  
 Y 350 IRIKMSKQIKKYNVNRV 369  
 D 829 VFAI-LAEIRHINIVYVNV 847

RESULT 13  
 HSB2: GRYSA  
 ID HSB2: GRYSA  
 AC p33126;  
 DT 01-OCT-1993 (rel. 27; Created)  
 DT 01-OCT-1993 (rel. 27; Last sequence update)  
 DT 01-NOV-1995 (rel. 42; Last annotation update)  
 DE BEAT SHOCK PROTEIN B2.  
 GN HSPB2.  
 OS Oryza sativa (rice).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzoae;  
 OC Oryza.  
 OX NCBI Taxid 4540;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN CV. TAICHUNG NATIVE ONE;  
 RA MEDLINE: 94198757; PubMed: 7764644;  
 RA van Brusegom F., de Koster R., Garcia A., Claus B., Glendon J.,  
 RA "Heat inducible rice hspB2 and hsp70 are not always co-regulated".



KL Plantia 193:57-66(1994).  
CC -1- FUNCTION: MOLECULAR CHAPERONE. HAS ATPASE ACTIVITY  
CC (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).  
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb.sib.ch/announce/>  
CC or send an email to [license@isb.sib.ch](mailto:license@isb.sib.ch)).  
CC -----  
DR EMBL: Z11920; CAA77978.1; -  
DR PIR: S25541; S25541.  
DR HSSP: P07900; 1YER.  
DR InterPro: IPR001404; -  
DR Pfam: PF00183; HSP90.1.  
DR PRINTS: PR00779; HEATSHOCK90.  
DR PROSITE: PS00248; HSP90.1.  
KW Chaperone; ATP-binding; Heat shock.  
SO SEQUENCE 699 AA; 80194 MW; 5381D14B572B3E CRC64;  
  
Query Match 4.4%; Score 89; DB 1; Length 699;  
Best local Similarity 20.4%; Pred. No. 12;  
Matches 64; Conservative 49; Mismatches 98; Indels 102; Gaps 17;  
  
QY 23 EYHSPEGLIFWNPSPKPEVSGSSGSSGKQKMWPPHNNQYQEDHDVYQKQKEDKS 62  
DB 201 EPISTYPLS-LMIETKTEKTSDEDEEEK-----KDEKSKVDVDVDEKKEKKEKK 251  
QY 83 KLK-LSDMF-----NP-FKPEEVVMTDMKAPVWEGYVRAVLDVYAKSKITVGLT 133  
DB 252 KIKVSHHMLVANKQPIWKRKPELITKEVAA-----FYSLIINMW----- 293  
QY 134 VPAVKRIEHTLEFELTSANKHMGHRYFYVWVDVSKMPL-----TELGL 182  
DB 294 -----EEHL-AVKHESVQLEFFKAVLLVPRKAPDLDITRKKLNNIKLYR 339  
QY 183 KSP-----KVEVLPKRRKQOVSMVKMTIGEHIVAHIGRE-----VD 220  
DB 340 KVTIMNGCELLE-W--LSFYKGLVDSHDLPLNLSRELQGNKLIKLYRKLVKVCVE 395  
QY 221 FLFCM-----DVDQVFD-----EFQV---ETLGSVAULQAWMYADPEDEYERKES 267  
DB 396 LFEFLANRKKDYKRYEAFSKNLKLGTHHDSIRNKTAFGLRYHSTKSGDEL--SLQDV 453  
QY 268 AVYIPFEGCGPY 280  
DB 454 VTKMKRQNDIY 466  
  
RESULT 14  
DMPN\_PSES SP STANDARD; PRT; 517 AA.  
ID DMPN\_PSES SP STANDARD; PRT; 517 AA.  
AC P19732;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE PHENOL HYDROXYLASE P3 PROTEIN (PC 1.14.13.7) (PHENOL 2-MONOXYGENASE  
DE P3 COMPONENT).  
GN DMPN OR PHENOL.  
OS Pseudomonas sp. (strain CF600).  
OG Plasmid pV1150.  
OC Bacteria; Proteobacteria.  
OX NCBI\_LaXid=106;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91072230; PubMed=2254258;  
RA Nordlund I., Powlowski J., Shingler V.;  
KT \*Complete nucleotide sequence and polypeptide analysis of

HT multicomponent phenol hydroxylase from pseudomonas sp. strain  
RT CF600.?  
KL J. Bacteriol. 172:6826-6833(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BH:  
RA Takeo M., Maeda Y., Okada H., Miyama K., Mori K., Ito M.,  
RA Fujita M.;  
KL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: CATABOLIZES PHENOL, AND SOME OF ITS METHYLATED  
CC DERIVATIVES. P3 IS REQUIRED FOR GROWTH ON PHENOL, AND FOR  
CC IN VITRO PHENOL HYDROXYLASE ACTIVITY.  
CC -1- CATALYTIC ACTIVITY: PHENOL + NADH + O(2) -> CATECHOL + NADH(+) + H(2)O.  
CC -1- COFACTOR: FAD, AND REQUIRES FE(2+) FOR ACTIVITY.  
CC -1- PATHWAY: FIRST STEP OF PHENOL BIODEGRADATION.  
CC -1- SUBUNIT: THE MULTICOMPONENT ENZYME PHENOL HYDROXYLASE IS FORMED  
CC BY P0, P1, P2, P3, P4 AND P5 POLYPEPTIDES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb.sib.ch/announce/>  
CC or send an email to [license@isb.sib.ch](mailto:license@isb.sib.ch)).  
CC -----  
DR EMBL: M60276; AAA25942.1; -  
DR PIR: D28864; BAA06017.1; -  
DR PIR: D37831; D37831.  
KW Aromatic hydrocarbons catabolism; oxidoreductase; monooxygenase;  
KW NADP; Flavo-protein; FAD; Iron; Plasmid.  
SO SEQUENCE 517 AA; 60522 MW; 4EED4B40ED74F9C CRC64;  
  
Query Match 4.4%; Score 88.5; DB 1; Length 517;  
Best local Similarity 18.9%; Pred. No. 8.6;  
Matches 59; Conservative 42; Mismatches 128; Indels 84; Gaps 12;  
  
QY 72 EKKQKREKSKSLKIDMNPYKREVVVMTDMKAPVWEGYVRAVLDVYAKQ----- 126  
DB 25 FTLYQKKED-----VPLPHFGKLTIDMK-----WDDP-KLIMTYKYLAEKK 71  
QY 127 KITVGLTVAVGKRYLPHYEELTSANKHMGHRYFYVWVDVSKMPLITGLPSEK 186  
DB 72 KYAIFDAFAQNNQHONISDARYNALKFLTAVSPLEVFQAGSRAVG-----RQFS 124  
QY 187 VPEVKRRKQOVSMVKMTIGEHIVAHIGRVDPL--FGMVQVQVPOHPGVELLIGESV 244  
DB 125 GAGARACQMOAIDELRHVQTVVHANSHYKHHIDGLDPAIMYDRV----- 170  
QY 245 AOLQAWMYKADPD-----EFTYRKRESAAYIPFGCGPY-----YHAA 283  
DB 171 -----WYLSVRSYMDQARTAGPFFLLTAVSPREYVLTNLLFVPRMGAAVNDMAIV 224  
QY 284 IFG-----GIPTVOLNITQCFKGLDKKKNDE--AQMDHSHLNNKPYLN----- 328  
DB 225 TFGFSQSDSEARHMTLGLVEIKFMLEOHEDNVPILQRMWDKFWKRYRLLITIGMMQY 284  
QY 329 KPYKILSPYCW 340  
DB 285 LPKKVMSEBAM 296  
  
RESULT 15  
VHTU\_BPT7 SP STANDARD; PRT; 536 AA.  
ID VHTU\_BPT7 SP STANDARD; PRT; 536 AA.  
AC P03728;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 01-FEB-1991 (Rel. 17, Last annotation update)  
DE HEAD-TO-TAIL JOINING PROTEIN.  
GN 8.

